

Real-time signature embedding in video

Field of the Invention

This invention relates in general to the field of signal authentication and more particularly to the embedding of signatures in an audio-visual signal for authentication of images and video.

5

Background of the Invention

The success of digital imaging and video has lead to a wide use of this technology in many fields of everyday life. Technology to edit, alter or modify digital images or video sequences is commercially available and allows modifications of the contents of said images or videos without leaving traces. For a variety of applications, such as evidential imaging in law enforcement e.g. from security cameras, medical documentation, damage assessment for insurance purposes, etc., it is necessary to ensure that an image or video has not been modified and is congruent with the image or video originally taken. This led to the development of signal authentication systems for which an example is shown in Fig. 1, wherein a signature is created at 1.20 for an audio-visual signal, such as an image or video, which is acquired in 1.10. The signature is embedded e.g. as a watermark in 1.30 into the signal. Thereafter the signal is processed or tampered in 1.40, played, recorded or extracted in 1.50 and finally verified in 1.60 in order to either ensure that the authenticity of the signal is proven or that modifications of the signal are revealed.

10
15
20

Embedding data into a video-signal is known from US-B-6 211 919 wherein an analogue video signal is converted to a digital video signal into which data is embedded and then converted back to an analogue video signal. Error correction across frames is implemented in order to compensate for transmission losses. The solution disclosed therein is of complex technical nature requiring large buffer memories for storing the entire frame or several frames of the video signal. These memories are expensive and it is therefore desired to minimise the amount of memory needed.

25

Furthermore, especially for the above mentioned applications of authenticating signatures, it is important that each video frame possesses the capability to authenticate itself, because in e.g. the above mentioned security camera application, not all frames of a sequence

are stored, e.g. only every fiftieth frame, likewise for medical imaging, perhaps only a subset of images are retained. In general it is not known which frame will be recorded and which will be discarded. Consequently, all information required to authenticate a certain frame of a video sequence must be available in and derivable from the frame itself. This is not possible, when a frame has a dependency on preceding or subsequent frames, as in the above document, in order to enable authentication of the frame.

The signature calculation and embedding has to take place as soon as possible after the generation of the video signal in order to prevent the video being tampered before authentication information is stored in it. Therefore it is an advantage if the signature calculation and embedding is placed close to the image capturing device, e.g. inside a security camera, and the signature calculation and embedding takes place in real-time on the video stream generated. Today's solutions, as disclosed in the above document, are technically complicated and expensive.

Finally, according to the prior art, in order to embedded the signature bits calculated in 1.20 for an audio-visual signal, such as a digital image, inside the audio-visual signal itself as a watermark in 1.30, an entire frame of the audio-visual signal has to be buffered in a large, expensive memory while the signature bits for the frame of said audio-visual signal are calculated, the watermark having the signature bits as a payload is constructed, and finally said watermark is embedded inside said frame of the audio-visual signal. This renders such solutions expensive due to the amount of expensive memory needed.

Thus, the problem to be solved by the invention is defined as how to provide low-cost real-time generation of an audio-visual signal with self-authenticating frames.

Summary of the Invention

The present invention overcomes the above-identified deficiencies in the art and solves the above problem by embedding a signature in an audio-visual signal, such as a video signal or a digital image, in a way that completely obviates the need to buffer an entire frame of the audio-visual signal in a large memory while the signature bits are calculated and the watermark is embedded, thus dramatically reducing the cost of the memory needed, according to the appended independent claims.

According to embodiments of the invention, a method, an apparatus, and a computer-readable medium for authenticating an audio-visual signal are disclosed.

According to these embodiments, a signature is formed based on a first portion of a frame

of said audio-visual signal. Thereafter the signature formed is embedded in said audio-visual signal in said first portion or at least in a second portion of the frame to be authenticated whereby said portions are patterns of horizontal lines of said audio-visual signal and have fewer lines than the total number of lines of the entire audio-visual signal.

5 Thus a real-time low-cost solution, is proposed needing only memory for some lines of the audio-visual signal instead for memories storing entire frames of the audio-visual signal. All information required to authenticate the frame is put into the frame itself, rendering each frame self-authenticating.

10 These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereafter.

Brief Description of the Drawings

Preferred embodiments of the present invention will be described in the following detailed disclosure, reference being made to the accompanying drawings, in which

15 Fig. 1 shows a Prior Art authentication system;
 Fig. 2 shows an embodiment of the invention;
 Fig. 3 shows another embodiment of the invention;
 Fig. 4 shows a further embodiment of the invention;
 Fig. 5 illustrates an apparatus according to another embodiment of the
20 invention; and

 Fig. 6 illustrates a computer readable medium according to still another embodiment of the invention.

Description of preferred embodiments

25 A video signal, although representing a 2D image, is transmitted and handled as a one-dimensional signal by scanning the image line by line. Analogue or digital video is classified into interlaced and non-interlaced, also called progressive scan, video. For example video signals according to the NTSC, PAL and SECAM standard are interlaced and most PC displays are non-interlaced, whereas HDTV (High Definition Television) signals can be
30 either interlaced in higher resolution modes or non-interlaced in lower resolution modes.

Interlaced audio-visual signals, such as video, are defined in that each frame of said signals consists of two fields, whereby each field is a particular division of said frame and contains every other horizontal line in the frame. When handling an interlaced video by e.g. transmitting or displaying it, the field containing all the odd lines, inclusive the topmost

scan line, is handled first and called the upper field; the field containing the even lines is called the lower field and is handled consecutively to create a single frame or complete image. Thus, for an interlaced signal lines 1, 3, 5, ... (i.e. all of the first field) are handled first, then lines 2, 4, 6, ... (i.e. all of the second field) are handled. Each field can be subdivided into segments of consecutive lines of said frame, so called slices, e.g. slices of three lines: [1, 3, 5], [7, 9, 11], [2, 4, 6] or [8, 10, 12]. A special case of slices of consecutive lines in an interlaced signal is when the first slice comprises all odd or even lines of a frame and the other slice the remaining even or odd lines of the frame.

Non-interlaced video displays each line of a frame in order, whereby a frame is defined as a complete image in a sequence of images constructing a video. Thus, for a non-interlaced signal lines 1, 2, 3, ... (i.e. all lines of the frame) are handled. Such a frame can be subdivided into slices of consecutive lines e.g. slices of three lines: [1, 2, 3] or [4, 5, 6].

Interlaced and Non-Interlaced video refers to capturing, transmitting and displaying video sequences.

A portion of a frame is defined as an individual share of said frame being part of said frame, e.g. a slice or a field as defined above.

A region of a frame of an audio-visual signal, such as a digital image in a video stream, is defined as a spatial region within said frame, e.g. the top, the centre, the bottom.

Fig. 2 shows an embodiment of the invention, wherein an audio-visual signal, captured in step 2.10 is interlaced. The upper field of a frame in the interlaced audio-visual signal, e.g. consisting of n lines, is assigned to a first portion, loaded and held in a memory circuit in step 2.20. A signature of the first field is calculated in step 2.30, whereby said signature comprises information for authenticating all regions of the frame as the first field contains all image content, albeit only alternating lines thereof. Subsequently the lower field of the same frame in the audio-visual signal, e.g. consisting of m lines, is assigned to a second field in step 2.40 and the second field is saved in the same memory circuit, replacing the first field in the memory circuit. Said memory circuit needs therefore to have maximally a capacity of m respectively n lines, preferably m or n line memories. Thus the memory requirements are limited to half the requirements of the prior art as discussed above. The signature bits of said signature need also to be saved for the next step 2.50, where the signature is embedded in the second field of said audio-visual signal being in said memory circuit. However, storage capacity for said signature bits is negligible compared with that required for storing pixels in an audio-visual signal such as video. The signature bits can e.g.

be saved in the n -th line of memory as in practice the second field often comprises one line less than the first field, i.e. $m = (n - 1)$, depending on the frame size.

Fig. 3 illustrates another embodiment in which the audio-visual signal is non-interlaced, captured by progressive scanning in step 3.10. A slice of said audio-visual signal consisting of N horizontal lines is loaded into and hold in a memory circuit of sufficient capacity for said N lines, such as N line memories, in step 3.20. Then the signature is calculated for said slice in 3.30. In case the current signature is to be embedded in the current slice itself, step 3.50 will follow directly. In case the signature is to be embedded in the next consecutive slice, the next slice is now loaded into the N line memories, replacing the current slice. If the current slice is already the last slice in said frame, the signature can only be embedded in the current slice itself. In case, a common signature for all slices is to be embedded, the signature for the current slice is added to a common signature with respectively previously calculated slices' signatures in optional step 3.50. If the current signature is only to be embedded in the slice currently in the N line memories, it is not combined with previously calculated signatures. The signature is embedded into the slice currently in the N line memories in step 3.60. Subsequently the audio-visual signal is either further processed, e.g. by storing or transmitting, if signatures for all regions of the image have been calculated, i.e. signatures for all slices have been calculated, or the next slice is loaded into memory by returning to step 3.20, alternatively, if a new slice has been loaded into the N line memories in step 3.40, the signature is directly calculated in step 3.30, and so on. Storage of the signature bits calculated is similar to that described in the previous embodiment. This embodiment requires only holding a slice in memory therefore requires N line memories. While a particular slice is in memory it is possible to calculate the signature bits for that slice, and embed the signature into that slice, preferably as a watermark. The watermark can carry a payload consisting of the signature bits for the slice itself, plus any preceding slices' signature bits. Thus the i -th slice can be embedded with signature bits from slices 1 to i . The first slice can only be embedded with the signature bits of the first slice, and the last slice can be embedded with any or all of the signature bits from the entire frame of said audio-visual signal. Thus, the signature bits of the first slice may be embedded into any slice, preferably all slices, whilst the signature bits of the last slice are only embedded into the last slice itself. Thus self-authentication of the image is maintained.

In Fig. 4 a further embodiment of the invention is shown. An audio-visual signal is captured in step 4.10. As mentioned above, audio-visual signals are captured by scanning lines which have a certain position within a frame forming an image in a sequence

of images/frames. In the current embodiment it is not distinguished between interlaced or non-interlaced signals. In step 4.20 the DC-value is calculated for the current line of said audio-visual signal and in step 4.30 signature bits are formed based on said DC-value of the current line. The signature bits calculated are either directly embedded in the current line itself in step 4.50 and calculation continues with the next line until signatures are calculated and embedded in all lines or the signature bits currently calculated are saved in memory in step 4.40 for later embedding in a subsequent line together with the signature bits for subsequent line(s) or the current signature bits, even in combination with signature bits calculated for previous lines, are both embedded in the current line and saved in memory for subsequent use. Thus, for inexpensive real-time operation a signature calculation scheme is shown which requires only a line memory, rather than the storage of an entire field as described in the first embodiment. The signature bit representing a given image area is calculated only from that area itself, and other nearby areas, which means one or some lines of the audio-visual signal treated by the invention. On top of this, the signature is based upon some image property, such as DC value, edges, moments, or histograms, which only requires computation and storing in memory of the property, not of the pixels. The memory requirements for calculating the signature are thus typically much less than a field memory, some line memories as in the above embodiment are sufficient, in certain cases even less memory is required, depending upon the property used. For example, calculating DC values is done by averaging, i.e. adding up the values of pixels of the audio-visual signal. In this case it is not necessary to store the pixel values themselves and the memory requirements are further minimised compared to the previous embodiments. Similarly for calculation of the watermark, once the payload is known, i.e. the complete signature is available, then forming the watermark can typically be done using only a few line memories because adapting the watermark to the image, in order to get the best trade between robustness and invisibility, involves looking at the image complexity, i.e. characteristics such as the amount of image activity in edges, texture etc., in localised areas around the watermarked pixel currently being calculated. This requires just a few line memories to hold the image pixels close by and the same line memories from above when in forming signature bits are used and no further memory circuits are necessary.

Fig. 5 illustrates an embodiment of the invention in a system 100 for authenticating an audio-visual signal. An audio-visual signal is generated in 110. Preferably the audio-visual signal is captured in 110 by an image capturing device camera, such as a surveillance camera or a CCD array and/or an appropriate means for capturing the audio

signal, such as a microphone. However, the audio-visual signal may also originate from a transmission signal, such as a video signal, or from a storage device, such as a harddisk drive or similar computer readable medium. The audio-visual signal is further processed in the apparatus 101 according to an embodiment of the invention. The audio-visual signal captured in 110 is fed into the apparatus 101. A slice of N lines of said audio-visual signal are stored and hold in memory 120. Memory 120 is built of N line-memories and comprises an additional memory for storing signature bits. The number N of lines is much lower than that of the entire audio-visual signal, an example is 3 line memories in means 120 for 480 horizontal lines in an audio-visual signal captured in 110. The extra memory needed for said signature bits is much lower than that for said lines, according to the discussion in the sections above. Means 130 communicates with said memory circuit 120 and calculates a signature for the lines in memory 120. The signature formed is based on the contents of the lines in memory 120. When the signature is formed, it is embedded in the lines still held in memory 120. The signature bits generated are saved in memory 120 for later use, such as embedding in subsequent slices of said audio-visual signal. The signature calculated is preferably embedded as a watermark, preferably a robust watermark, by means 140. A robust watermark is a watermark which is embedded in the audio-visual signal and which is not influenced by allowable image operations such as lossy compression. Subsequently said lines of said audio-visual signal with the signature embedded are fed out of apparatus 101 for further processing in 150. Subsequently the next N lines of the same frame of said audio-visual signal are loaded into memory 120, the signature is formed for the new line contents and embedded into the lines, preferably in combination with the signature bits previously calculated and saved in memory. The combined signature is also calculated by means 120. The above procedure is repeated until a signature has been calculated and embedded for all lines of a frame. Then memory contents in 120 is erased and a new frame generated in 110 is treated in 101.

Apparatus 101 is preferably implemented in the system 100 as a module, preferably comprising a microprocessor or similar electronic device such as a programmable array or similar electronic circuit.

Fig. 6 illustrates another embodiment of the invention comprising a computer readable medium 220 in a system 200 for authenticating an audio-visual signal whereby an audio-visual signal is generated in 230. Preferably the audio-visual signal is captured in 230 by an image capturing device camera, such as a surveillance camera or a CCD array and/or an appropriate means for capturing the audio signal, such as a microphone. However, the

audio-visual signal may also originate from a transmission signal, such as a video signal, or from a storage device, such as a harddisk drive or similar computer readable medium. A first program module 240 directs a computer 210 to form a signature for a slice of N lines of a frame of said audio-visual signal. In a second program module 250 said signature generated by the first program module is embedded in said slice of said frame of audio-visual signal, preferably as a watermark, more preferably as a robust watermark. The steps performed by program modules 240 and 250 are repeated with subsequent slices of lines of said frame until a signature has been calculated and embedded for the entire frame. Subsequently the audio-visual signal with the signature embedded is further processed, e.g. for authentication of the audio-visual signal, in 270.

In some applications of the invention, such as security imaging, only one of a plurality of frames, e.g. one frame in every 50 frames, is stored. It is therefore important that each frame is capable of authenticating itself without reference to preceding or subsequent frames. According to the invention the signature is embedded in the frame itself. The above method meets therefore this requirement as it treats each video frame as a separate still image. This also means that the method is equally applicable to both still images and video.

For security reasons, the signature calculation and embedding is placed as close as possible to the image capture device. This prevents the possibility of the audio-visual signal being tampered before the signature is calculated. Consequently the signature calculation and subsequent embedding, preferably as a watermark, preferably take place in real-time on the video stream generated inside an image-capturing device such as a camera. According to the invention, only a part of a whole frame of the video stream is stored in a memory. Therefore the method and apparatus according to the invention are well suited for real-time embedding of a signature. The person skilled in the art of signatures will therefore clearly use a type of signature generation which is adapted for real-time applications. However, the invention is not limited to a specific type of signature calculation.

In order to judge the authenticity of an image, a similar procedure to the signature formation is used, i.e. a signature is again calculated from a first portion of a frame of an audio-visual signal. In order to authenticate the contents of said portion, the original signature embedded in a portion of said frame is extracted and compared to the signature anew calculated for said portion, whereby the portion having embedded the original signature is not necessarily the same portion as for which the signature was originally calculated, e.g. the signature for lines 1,3,5 of a frame can be embedded in lines 13,15,17. Tampering is detected when the two signatures differ from each other. In case tampering is detected, an

analysis of the modification is undertaken, if it is desired to e.g. localise where in the contents of said frame tampering has occurred, depending on the information derivable from the signature embedded.

Applications and use of the above described signal authentication according to

5 the invention are various and include exemplary fields such as

security cameras or surveillance cameras, such as for law enforcement,
evidential imaging or fingerprints,

health care systems such as telemedicine systems, medical scanners, and
patient documentation,

10 insurance documentation applications such as car insurance, property
insurance and health insurance.

The present invention has been described above with reference to specific
embodiments. However, other embodiments than the preferred above are equally possible
within the scope of the appended claims, e.g. different field patterns than those described
15 above, performing the above method by hardware or software, combining features from the
embodiments such as e.g. forming slices within fields for interlaced content of audio-visual
signals, or embedding signatures in interlaced content using some line memories, etc.

Furthermore, the term "comprising" does not exclude other elements or steps,
the terms "a" and "an" do not exclude a plurality and a single processor or other unit may

20 fulfil the functions of several of the units or circuits recited in the claims.